



Autonomous ultrasonic inspection using Bayesian optimisation and robust outlier analysis



R. Fuentes^a, P. Gardner^a, C. Mineo^b, T.J. Rogers^a, S.G. Pierce^b, K. Worden^a, N. Dervilis^a, E.J. Cross^{a,*}

^a Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

^b Centre for Ultrasonic Engineering, University of Strathclyde, 204 George Street, Glasgow G1 5PJ, UK

ARTICLE INFO

Article history:

Received 29 August 2018

Received in revised form 28 January 2020

Accepted 11 April 2020

Keywords:

Non-destructive testing (NDT)

Ultrasound

Gaussian process (GP) regression

Bayesian optimisation

Outlier analysis

ABSTRACT

The use of robotics is beginning to play a key role in automating the data collection process in Non Destructive Testing (NDT). Increasing the use of automation quickly leads to the gathering of large quantities of data, which makes it inefficient, perhaps even infeasible, for a human to parse the information contained in them. This paper presents a solution to this problem by making the process of NDT data acquisition an *autonomous* one as opposed to an automatic one. In order to achieve this, the robotic data acquisition task is treated as an optimisation problem, where one seeks to find locations with the highest indication of damage. The resulting algorithm combines damage detection technology from the field of data-driven Structural Health Monitoring (SHM) with novel ideas in uncertainty quantification which enable the optimisation routine to be probabilistic. The algorithm is sequential; a decision is made at every iteration regarding the next optimal physical location for making an observation. This is achieved by modelling a two-dimensional field of novelty indices across a part/structure which is derived from a robust outlier analysis procedure. The value of this *autonomous* approach is that the output is not only measured data, but the most desirable information from an NDT inspection – *the probability that a component contains damage*. Furthermore, the algorithm also minimises the number of observations required, thus minimising the time and cost of data gathering.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non Destructive Testing (NDT) is an integral part of the assessment of quality and structural integrity in engineering components in-service and post-manufacture. Currently, most NDT is carried out manually, however, recent increased attention to robotic-based inspection [1] promises to change matters. This use of robotics is a major boon to inspecting high-value and/or safety-critical assets; it also removes humans from harm when inspections are needed in harsh and dangerous environments.

Whilst the motivations for robotic NDT are clear, and research is under-way [1,2], little or nothing has been done in the way of performing *autonomous*, as opposed to *automated*, inspections. In an automated inspection scheme, the robot path is programmed prior to the inspection, and the robot makes no decisions regarding which areas of a component it should prioritise. The data collected has to be reported back to a human, or to further post-processing algorithms in order to assess the

* Corresponding author.

E-mail address: e.j.cross@sheffield.ac.uk (E.J. Cross).

state of the component. In a truly *autonomous* scheme, a robot would make its own decisions regarding the path it should take, and should also continuously perform its own critical assessment of the component: does it contain a defect or not, and with what probability? Has enough area been inspected, and are there critical areas that need more attention? The purpose of this paper is to formulate an algorithmic framework for such autonomous inspection.

This paper will focus on the particular use of ultrasound-based NDT, due to its current widespread use within engineering industry. However, the reader will recognise that the ideas presented are applicable to other types of testing. The proposed algorithm is demonstrated here using an example dataset originating from robotic inspection of a carbon-fibre wing panel with two known areas of delamination, as described in detail in [1].

The rest of this introduction will provide a general overview of the algorithm, with the following sub-sections providing a general background to ultrasound-based NDT, the challenges of autonomous inspection, as well as spatial sampling, which is closely related to the algorithm presented here.

1.1. Algorithm overview

The first question to be asked is, what is the *objective* of the robotic inspection? Here, this objective is defined as: determining, with high confidence, whether the component being inspected contains a defect, flaw or damage of a given minimum target size, ℓ and with the fewest possible number of measurements. Given such a clear objective, it is possible to cast the autonomous inspection problem as an optimisation problem. After an observation is taken, the optimiser should decide at what position the next observation should be taken next, so as to maximise the given objective. These decisions are thus carried out sequentially as information is being gathered. One method that deals particularly well with this sequential design problem is *Bayesian optimisation*, which will be discussed in detail in Section 2. The general idea is to fit a Bayesian nonlinear regression model – such as a Gaussian process (GP) – to the objective function. This allows one to infer the objective at unobserved locations. As GP regression is a Bayesian method, it can not only predict the objective at unobserved locations, but also estimate the uncertainty in the prediction. This allows one to place observations in areas of both *high uncertainty* and with *high values of the objective function*. This strategy maximises the information gain of every single observation that is taken, it can be carried out sequentially and in an autonomous fashion.

For optimisation to work, a good objective function is required. Here, the objective of the problem is to detect defects, so the objective should capture information regarding whether a particular area contains a defect or not. This task can be cast as a problem of *novelty detection* [3]; the basic idea being to model the probability density of data known to belong to an undamaged class, and to then compute a distance between *new* observations and the mean of that density. Here, the data are assumed to be Gaussian, and the *Mahalanobis squared-distance* (MSD) will be used to provide a *novelty index*. The MSD deals well with multivariate data, and is thus useful in the context of analysing NDT data. If the highest values of a novelty index are found, these correspond to the most abnormal observations with respect to the rest. One of the key novelties of this study will be the use of novelty indices as the objective function to be optimised.

Combining Bayesian optimisation, with a novelty index seems like a good idea; however, the novelty detection must be carried out both on-line (since it must guide a robot) and must also, when capturing a reference undamaged set, ignore abnormal observations. If outlying observations are taken into account for training of the reference set, the novelty detector will become biased towards them and subsequently mask potential damage sites. A remedy to this problem is to use *robust statistics*, and in particular, the *Minimum Covariance Determinant* (MCD): a robust method for estimating the parameters of a Gaussian distribution. The MCD is discussed in detail in Appendix A; it attempts to discriminate between outlying and inlying observations when outliers are present in the training set, and thus does not bias the reference set toward the outliers. This enables an autonomous inspection algorithm that is entirely data-driven – it does not use external information to form a model of a healthy component.

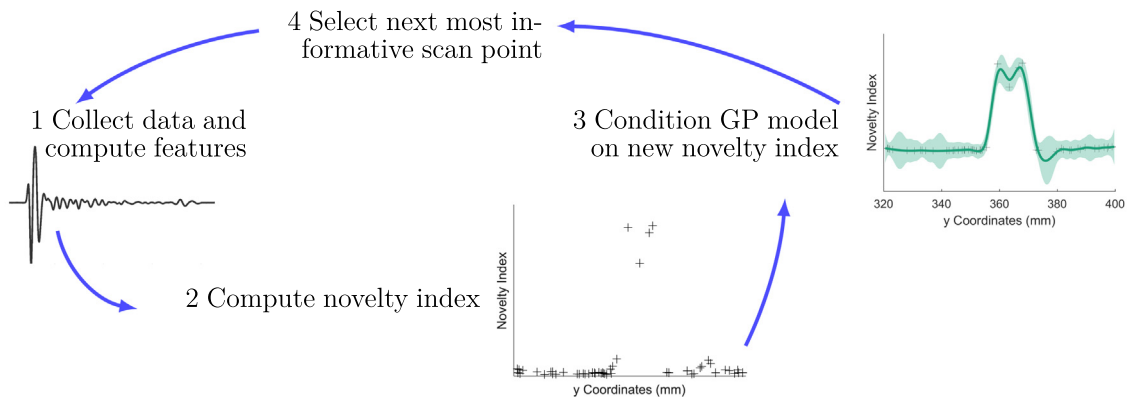


Fig. 1. Overview of the steps involved in the autonomous inspection algorithm.

An overview of the autonomous inspection scheme is illustrated in Fig. 1. Step one involves data collection and feature extraction from raw data. Step two involves modelling these features using robust outlier analysis in order to establish a threshold. Steps three and four form the Bayesian optimisation step, first modelling the objective function with a GP, and then choosing a location that maximises information gain.

A major assumption made in this autonomous inspection strategy is that the physical area being scanned is nominally homogeneous. It is assumed that variability in the material and geometric characteristics, together with any measurement/-computation noise, can be modelled by a Gaussian distribution. If a component did have two sections with different characteristics, this would add multiple modes to the underlying probability distribution of undamaged class features, and render the MSD ineffective. This issue could still be dealt with via a clustering analysis, and appropriate novelty indices exist that capture this situation well [4]. However, in this scenario of actively looking for data that are dissimilar, it would not be possible to distinguish between damage and a benign change in normal properties. The key here is to use *prior knowledge*. A good use of prior knowledge is to assume the robot knows which sections of a component are expected to be nominally the same. This can be achieved by performing independent inspections over areas that are known to be homogeneous in their material and geometrical properties. If a component is homogeneous, there is no problem. If not, one can elicit information from Computer-Aided Design (CAD) models, as already demonstrated in robotic path planning in NDT inspections [2]. In this paper, the emphasis is placed on achieving *autonomous inspection*, through the use of Bayesian optimisation and robust outlier analysis.

1.2. Ultrasound signal features

This paper will use example data from ultrasound-based NDT, due to its current widespread use in practice. This section will provide the reader with an understanding of the type of signals, and the features extracted from such signals, as well as the type of physical testing required to collect ultrasound data.

A typical ultrasound pulse is shown in Fig. 2a, with two time indices marked as t_a and t_b . These times correspond to reflections from the front and back wall of the inspected specimen respectively. A pulse of this kind effectively constitutes an A-scan. The important information is the time difference $t_f = t_b - t_a$, and this is often referred to as the ultrasonic Time of Flight (ToF). ToF can be related to the thickness of the specimen, if the propagation speed of bulk waves for the material is known. Another feature of interest is the ratio $x(t_a)/x(t_b)$ (where $x(t)$ is the measured amplitude of the ultrasound pulse), as this contains information about the attenuation of the wave as it travels through the thickness of the component. An A-scan thus gives information about a single physical coordinate on a surface.

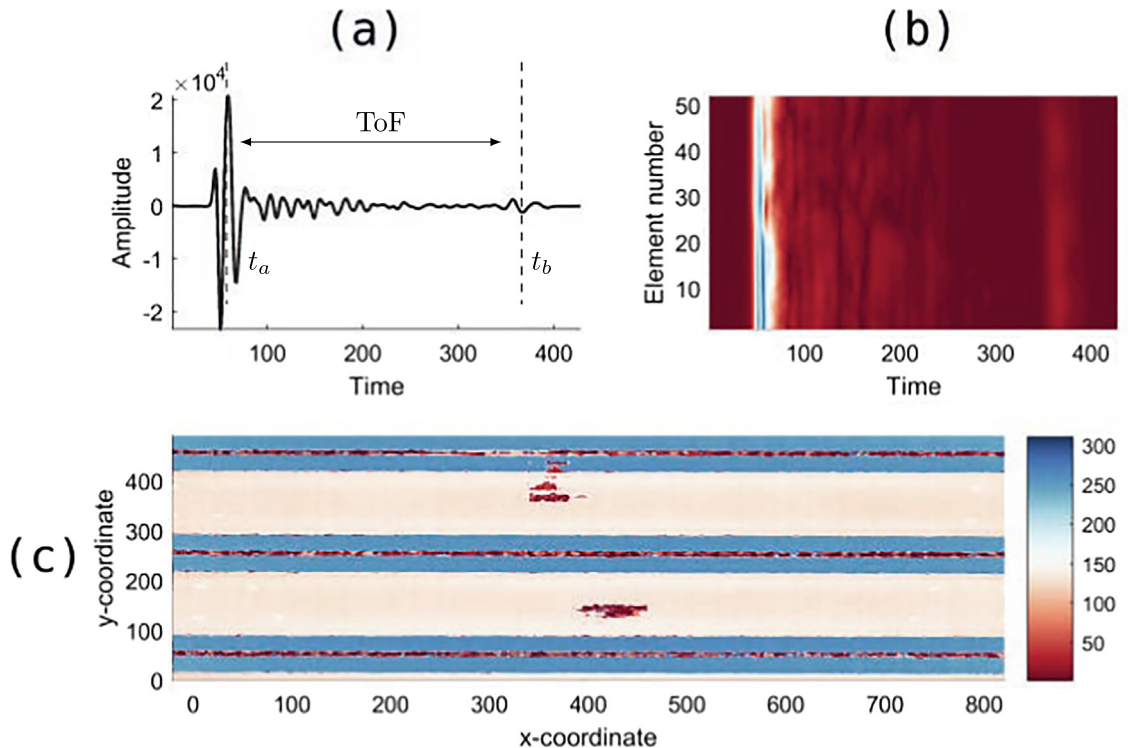


Fig. 2. Illustration of a) single ultrasound reflection (A-scan), b) reflections along a phased array probe (B-scan) and c) time of flight map along the two dimensions of a composite plate (C-scan).

A B-scan can be formed by collecting a series of A-scans along a line (illustrated in Fig. 2b), while a C-scan is formed by collecting a series of B-scans, to give a two-dimensional grid of ultrasound pulse information (illustrated in Fig. 2c). Note that higher times of flight in Fig. 2c imply greater plate thickness. The scanned component in Fig. 2c is a wing panel, the salient features visible are two visible regions of delamination, the stringers and the variable panel thickness. The stringers are evident by a low ToF index, due to their thickness being higher than the maximum value that can be captured within the collected data.

1.3. Towards practical autonomous inspection

In practice, A, B and C-scans are often performed manually, with the aid of engineering expertise, in order to carry out four key tasks:

1. Physically scan the structure with an ultrasound probe.
2. Interpret the results of the collected dataset and form a decision over the quality of the data, and the state of the component.
3. Decide which areas, in any, of the component need re-scanning.
4. Provide an assessment of the best course of action, and recommend an appropriate repair strategy.

Recently, there has been a move towards an increased use of robotics for automating the process of acquiring the necessary data, where ultrasound data acquisition hardware is combined with robotic interfaces in order to automate the physical scanning of the structure (task one in the hierarchy above). Examples of this can be found in [5–7] and more recently in [1,2], where problems such as path planning and integrating CAD geometries into the scanning process have been addressed. Other important issues, such as dealing with uncertainty in the positioning of autonomous robots, have been addressed in [8] using Bayesian filtering methods.

The interpretation of results, the second task in the hierarchy, has received a noteworthy amount of attention in the NDT research community. Research has evolved from investigations into extraction of useful features from the raw data [9–11], into the use of machine learning and novel statistical inference methods applied to those features in order to detect, cluster and classify defects [12–16].

The third task in the hierarchy involves deciding appropriate places to scan, based on information collected so far about the component. The continuous path-planning problem using optimisation has been investigated before in a general context [17], but significantly less research has been carried out towards this end in NDT. The state of the art in path planning for robotic ultrasound inspection, at the moment, is to use CAD models to plan the scanning trajectory, where research has focused around issues with the feedback control of the ultrasonic probe head [2].

The objective of performing a scan is to find defects that are larger than a minimum given size. The current solution in pulse-echo ultrasound measurements is to define a spatial grid with a high enough resolution to capture this minimum expected flaw size. Whilst this approach has worked so far, there are two major issues:

1. The scan can take a long time, and there are cost implications associated with this.
2. A large quantity of data needs to be stored and post-processed later.

These effects are exacerbated when large scanning areas are involved, as is typical in aerospace industries. Depending on the motivation for the inspection, the scan time may not present an issue; for high-value, safety-critical engineering components, large scanning times are acceptable. However, the data storage and processing requirements of a dense grid over a large area can make the problem intractable.

The data storage requirement can be alleviated by compressing the raw data for later post-processing, leveraging advanced signal processing techniques such as wavelet transforms [18], sparse signal modelling [19], and more recently compressive sensing [20].

1.4. Spatial sampling

Efforts in developing compression algorithms for time domain data will do little to address the problem of dealing with large quantities of spatial data. An important question is: what is the minimum number of observations required to capture the spatial statistics of the data collected from the component? This problem has been studied in a number of research fields, (see [21] for an overview). The most active research communities in this field are those of geo-statistical modelling, environmental modelling [22], and disease mapping [23], where the motivations for studying spatial sampling schemes arise from the fact that measurements can be sparse; this raises the need to interpolate data in regions with low observation density, while quantifying the uncertainty around those predictions.

In spatial statistics, the features of interest are treated as random variables across space; they can be separated into two categories, those which can be assumed identically and independently distributed (iid), and those that are autocorrelated across space. NDT data falls in the latter category, where values of NDT data features in homogenous materials will tend to be similar to those observed a small distance away (there is some experimental support for this [24]).

Modelling the spatial correlation of stochastic processes is central to the task of interpolation in spatial analysis [25]; the correlation quantifies the influence of a variable on itself at a distance r away. In spatial statistical analysis, this influence is modelled through the use of a covariance, or correlation function. The quality of spatial interpolation, based on correlations or covariates, relies heavily on the properties of the covariance function used; it should encode the overall smoothness of the random field being modelled, as well as the overall length-scale over which this random field correlates. Analysis of spatial variation has been of interest in NDT, and has been used as a tool to assess optimal resolution and scanning strategies [26,24].

When an application is risk-sensitive – which is the case in NDT data analysis – it is critical to use methods that are probabilistic. Gaussian Process (GP) regression is adopted here as a flexible tool for Bayesian, nonlinear and nonparametric regression; it will be discussed briefly in Appendix B. In previous work, GPs were used to replace more traditional empirical variograms when smoothing NDT data features across space. An interesting application of spatial modelling to outlier analysis is presented in [27] where the GP was modified to derive ‘extreme spatial processes.’ In this work, the GP is used to fit a surrogate model to novelty scores from a robust outlier analysis procedure; this approach is novel compared to other work in this area, where GPs are usually used to fit data features directly.

One important aspect of spatial variation and NDT data analysis is that of *Probability of Detection* (PoD); the likelihood that an object will be detected in a particular measurement scenario. Its origin can be traced back to the development of radar, but in the present day it is widely used to quantify the reliability of NDT processes [28]. PoD curves describe the probability that a particular NDT system will detect a flaw as a function of the flaw size. Traditionally, these curves are built experimentally, by performing blind scans of calibration pieces with flaws of various sizes, and quantifying the false negative rates [29]. PoD is affected by a number of factors: measurement noise, spatial scanning resolution and the scattering characteristics of the expected flaw with respect to the ultrasound wave. Evidently, the spatial resolution of a scan will affect the overall probability of detection. The problem of evaluating PoD curves from spatial random fields that exhibit autocorrelation has been investigated in [30]. This is interesting from the point of view of the current work, as it links the idea of using a predictive model to the computation of the more classical assessment of reliability of the NDT process based on PoD. While in [30], the surrogate being used is Polynomial Chaos model, this paper models spatial variation using a GP; the analysis of PoD, however, would follow similar lines to those of [30].

1.5. Contributions of this paper

The novelty of the work here stems from the combination of spatial statistics with novelty detection, which directly enables an optimised acquisition, in terms of minimum number of measurements, through the framework of Bayesian optimisation. The acquisition of data is sequential and performed fully autonomously, with the system deciding the next optimal point to scan, guided by estimates of uncertainty and high discordancy values.

1.6. Outline of the paper

As explained in Section 1.1, and illustrated in Fig. 1, the *autonomous inspection algorithm* consists of two main ingredients: 1) setting a robust threshold based on MCD, and 2) Bayesian optimisation to guide the next scan point. Following a short overview of some necessary theory, a detailed description of the proposed method is found in Section 4. Throughout the paper, illustrations of the various aspects of the procedure are drawn from ultrasound data collected from the aerospace panel discussed in Section 1.2. These provide subsets of the data illustrated in Fig. 2. Section 5 provides a more detailed experimental investigation of the autonomous inspection algorithm, using this data set. Finally, Section 7 will present some conclusions and directions for future work.

2. Bayesian optimisation

It is prudent to begin by discussing Bayesian optimisation and how such a scheme can be used to select appropriate scanning points, based on the features generated by the robust outlier analysis. Bayesian optimisation is used primarily in situations where it is expensive to evaluate the function being optimised, as is often the case in the field of Design of Experiments (DoE). The Bayesian optimisation technique presented is based upon the GP as a Bayesian nonlinear regression model. However, before delving into mathematical details, the motivation and reasoning for selecting this type of optimisation algorithm over others will be described first.

2.1. Reasoning and motivation

In a move toward autonomous inspection, the task becomes the selection of optimal points to scan, or ‘query’, in order to achieve an estimate of the condition of the component to within a specified confidence level. In this paper, this task is addressed as a problem of optimisation. As discussed in the introduction, the objective being modelled is the distribution of novelty indices across the specimen. These indices are calculated by means of a robust outlier analysis, the FastMCD described in Appendix A.

In general, the optimisation task is to find the arguments of a function that maximise (or minimise) an objective (or cost) function. If at the end of a scan, one analyses the data and finds that the maximum novelty score falls under a novelty threshold, then confidence is established in absence of anomalies. Alternatively, if a significant number of observations lie above the novelty threshold, this indicates a problem. The task of inspection is thus a search for regions that differ most from the rest, in showing the highest novelty scores. The problem of searching for outliers in a spatial domain can, therefore, be cast as an optimisation problem, where the objective, or cost function being maximised is the novelty score.

Optimisation is used throughout science and engineering, often to determine parameters in statistical or physics-based models. If the objective function of interest is differentiable, a popular technique is *gradient descent*, which takes steps in the steepest directions on the objective function surface. While this is often effective, it is not suitable for the problem here; there is no reason to believe that taking the direction of greatest change in the novelty score will lead to the location of greatest damage. The optimisation is, therefore, chosen to be a global and gradient-free method.

Because of the nature of inspection, it is desirable for the optimisation method to quantify uncertainty in its predictions; this would allow the user to tune the search according the acceptable risk in a given inspection problem. Bayesian techniques are well suited to this end. One example of a Bayesian approach has previously been investigated for ultrasound test parameter optimisation [31], a related problem to that of finding an efficient search path. The issues here are that the search must be sequential, and without the aid of a forward-model; Markov Chain Monte Carlo (MCMC) methods will not do well with these two requirements. An advantage of Bayesian approaches to optimisation, that fit a *surrogate model* to the objective function [32,33], is the reduced number of evaluations of the objective required. These methods are important when function evaluations require expensive simulations or experimental campaigns.

2.2. Bayesian optimisation Background

Bayesian optimisation can be cast as a sequential design problem: given all the observed data so far, \mathcal{D}_t ,¹ what is the “optimal” point to observe next, such that maximum information gain is achieved? The information gain is estimated directly from the surrogate model, by forming a posterior distribution over the objective function, denoted here as f , given the data observed so far \mathcal{D}_t . This includes both the coordinates \mathbf{x} and observations \mathbf{y} , so $\mathcal{D}_t = \{\mathbf{x}_t, \mathbf{y}_t\}$. This posterior can be written down as,

$$p(f|\mathcal{D}_t) \propto p(f)p(\mathcal{D}_t|f) \quad (1)$$

This distribution contains useful information about where to place the next observation, in terms of both areas of high uncertainty as well as areas of high values of f . The sequential search needs to balance between *exploration* and *exploitation*. Exploration involves searching in areas of high uncertainty, whilst exploitation involves placing observations in areas of high function values, in order to narrow down to the values of \mathbf{x} that lead to globally maximising f . The difference is between high-risk high pay-off steps and low-risk low pay-off. Bayesian optimisation balances exploration and exploitation through the use of an *acquisition function* that evaluates an expected utility $u(\mathbf{x})$. The approach adopted here is described in Algorithm 1; it iterates over evaluations of the posterior over the objective, $p(f|\mathcal{D}_t)$ in order to find points that maximise $u(\mathbf{x})$, and then acquire new data at those points.

Algorithm 1 Bayesian Optimisation Algorithm

```

for  $t = 1, 2, \dots$  do
   $\mathbf{x}_t \leftarrow \arg \max_{\mathbf{x}} u(\mathbf{x}|\mathcal{D}_{1:t-1})$     ▷Maximise utility
   $y_t \leftarrow f(\mathbf{x}_t) + \varepsilon$                 ▷Sample the objective
   $\mathcal{D}_t \leftarrow \{\mathcal{D}, (\mathbf{x}_t, y_t)\}$         ▷Include new observations
end for

```

The objective function being assessed is modelled by means of a standard Gaussian process regression, a short review of which can be found in [Appendix B](#). With regard to the choice of acquisition function this is a user choice in the algorithm, two popular options are the probability of improvement (PI) [34] and the expected improvement (EI) [35]. The probability of improvement seeks to maximise the likelihood of the next assessment of the objective function improving its value, EI assesses the expectation over an improvement function which can be shown to balance, better, exploration and exploitation in the optimisation routine [35]. For this reason, EI is chosen for this work.

3. Experimental setup

An 800 mm × 550 mm composite panel was scanned using a six-axis robotic head, with a water-coupled ultrasound probe. The probe consists of 64 transducers, each of which fires a 5 MHz tone burst, and also acts as a receiver. The resolution of the scan can be adjusted, but for these results, the speed of the probe was set to yield a spatial resolution of 0.8 mm in the

¹ The subscript t in \mathcal{D}_t , and other relevant variables is used to denote observations up to and including t .

direction of the probe travel. The C-scan shown in Fig. 2c was generated using this data set. Details of the experimental procedure have been published in [1], to which the reader is referred for further details.

In order to capture the range of different depths of this composite specimen, an acquisition time of $9.6 \mu\text{s}$ was used. This equates to 480 samples at a sample rate of 50 MHz. The original ultrasound pulses collected through the water-coupled probe are susceptible to misalignment with respect to the arrival time of the first burst (the front wall reflection); this is evident, for example, in the illustrative B-scan shown in Fig. 2b.

Two data features extracted from the ultrasound pulses were used: ToF and attenuation, between the front and back wall of the specimen. In order for the misalignment across B-scans not to affect the data features, the ToF and attenuation were both estimated using a Hilbert envelope of the autocorrelation of each pulse, where the location of the maximum lag specifies the ToF and the attenuation is extracted from the argument of the envelope at that time lag. Note that these computa-

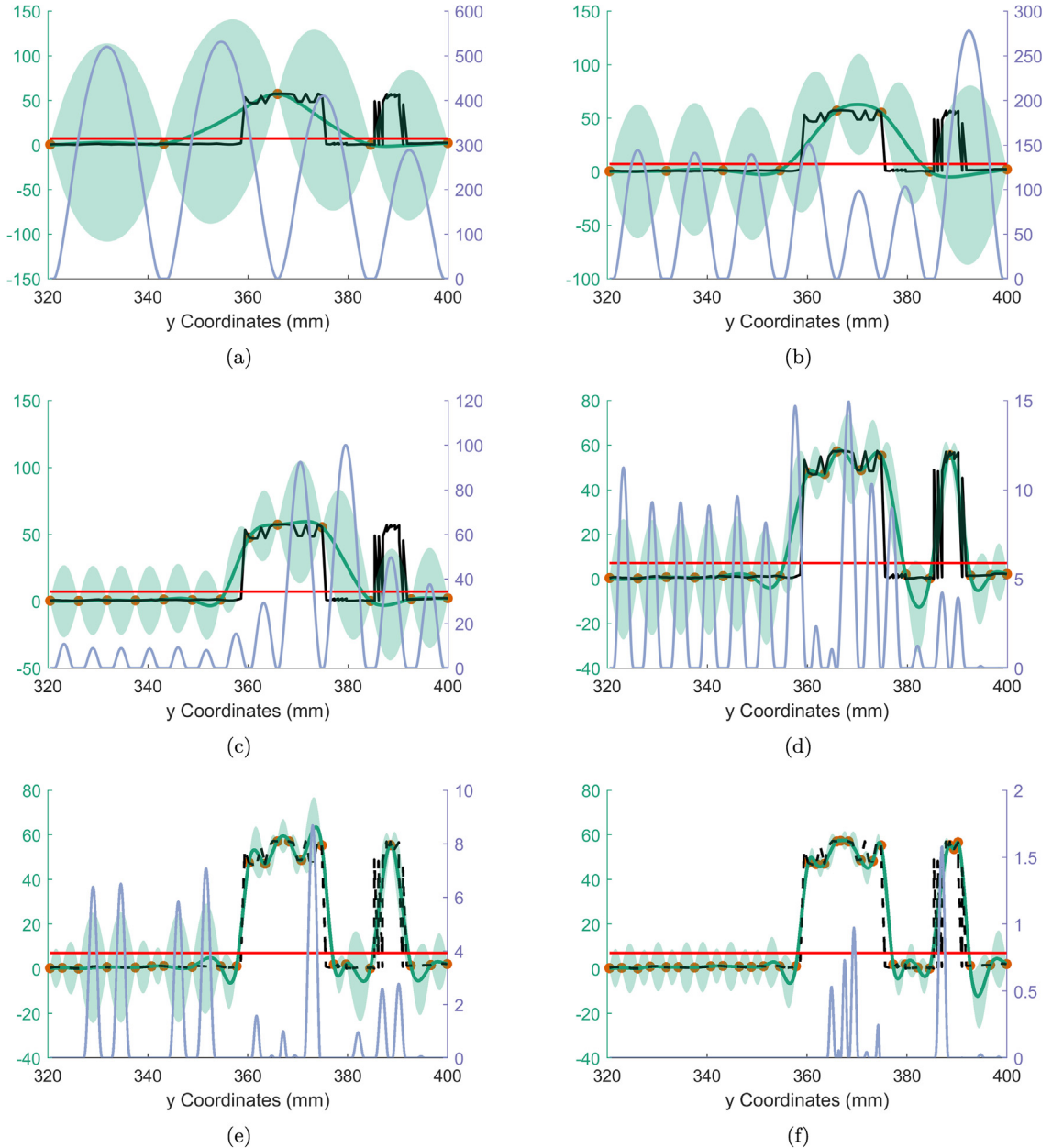


Fig. 3. Illustration of outlier search using Bayesian optimisation, on a one-dimensional slice of NDT data. The novelty scores (black) and its GP mean and 3σ -bounds (green), and acquired observations (orange circles) are shown on the left axis. The red horizontal line represents a 3σ threshold on the novelty scores. The expected improvement (blue) is overlaid and shown on the right axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tions could be performed directly in a compressed domain, thus requiring an even smaller quantity of data, adopting the techniques developed in [20].

4. Autonomous inspection

The proposed strategy for selecting scanning points in a sequential fashion is now discussed; this is presented as a Bayesian optimisation problem, where the objective function being maximised is the novelty index of the measured data. The novelty index describes the dissimilarity of a measured point against the rest of the group, and it is computed using robust measures to ensure that the reference model is not biased by potential outliers due to measurement noise or damage itself. The optimisation is performed over this novelty index to identify areas of interest. The result is a scan that minimises uncertainty across the spatial field while focussing on areas of potential damage. The goal is to achieve this with the least possible number of measurements. The implementation of the autonomous inspection algorithm involves important practical aspects such as initialisation, computation of damage thresholds and probability of damage.

The main iterations follow Fig. 1, where the four main elements of the iteration are:

1. Collect data and evaluate features.
2. Update robust mean and covariance including new observations and evaluate novelty indices for the entire set.
3. Condition the GP model on the new novelty indices.
4. Evaluate the Expected Improvement to find next best suitable location to gather data.

The overall search algorithm, incorporating the robust estimates of mean and covariance of NDT data features into the Bayesian optimisation framework is shown in Algorithm 2.² The feature vector \mathbf{y} could be any damage sensitive feature, like the ToF or attenuation mentioned previously; however, the MSD accommodates multivariate features, so high-dimensional objects like waveforms could work in principle [36]. In this respect, the overall algorithm presented is quite general.

An illustration of the application of Algorithm 2 is presented in Fig. 3. The one-dimensional section shown was chosen to cut across two regions of the panel containing damage. The example shows the result of the algorithm from several iterations, initialised with five random samples. Overlaid on the plots, in light blue, is the EI; its maximum indicates the point where the next observation is to be taken. Fig. 3 shows that, as expected, the algorithm balances exploration and exploitation. In the initial stage, it places more observations at high novelty index values, but then moves away to explore areas of low novelty index values but high uncertainty. As more observations are gathered and areas of low novelty are explored (and thus its variance reduced), the EI starts to weigh more heavily those locations with high function values. On the last instance shown in Fig. 3, observations have been placed evenly in areas of low novelty indices, and there is a higher density of observations placed around the areas of delamination. Also in this last instance, the EI indicates a clear priority for high novelty indices, meaning that if the algorithm were to run more iterations it would concentrate around areas of damage.

The mean and covariance of the data are re-estimated at every iteration of the algorithm in a robust manner using FastMCD. One should note that the MCD method defines inliers as those representing at least 50% of the data. These two points together imply that placing more than 50% of the observations around damaged regions will result in a shift where the damage area defines the inliers, and the data for rest of the undamaged component are defined as outliers. This outcome is of course undesirable; as a remedy, any observations flagged as outliers after the initialisation step must be excluded from the set used to re-estimate the mean and covariance of the data.

Algorithm 2 Autonomous Inspection

```

Initialise  $\mu$  and  $\Sigma$  and  $\mathbf{z}_{in}$  with low resolution C-scan
while  $\mathbf{v}_{t-1} \geq \alpha \sigma_n^2 \vee t \geq \text{budget}$  do
   $\mathbf{x}_t \leftarrow \arg \max_{\mathbf{x}} u(\mathbf{x} | \mathcal{D}_{t-1})$            ▷Maximise utility
   $\mathbf{z}_t \leftarrow \text{Acquire data at } \mathbf{x}_t$ 
   $\mathbf{y} \leftarrow \text{mahal}(\mathbf{z}_{in}, \mu, \Sigma)$            ▷Compute novelty scores
   $\mathbf{z}_{in} \leftarrow \mathbf{z}[\mathbf{y}_t < T]$                  ▷Update outlier-excluding set
   $\mu, \Sigma \leftarrow \text{FastMCD}(\mathbf{z}_{in})$          ▷Update robust statistics
   $\mathcal{D}_t \leftarrow \{\mathcal{D}, (\mathbf{x}_t, \mathbf{y}_t)\}$          ▷Include new observations
end while

```

4.1. Estimating probability of damage

An observation is flagged as abnormal if its novelty score is higher than the damage threshold. The GP provides a probabilistic estimate of the novelty scores over a two-dimensional spatial field. The probability that a given novelty score at spa-

² Note that in this case, the multi-dimensional features of raw data are denoted as the vector \mathbf{z} , whilst the novelty scores are denoted as y (and \mathbf{y} for a vector of scores), in order to keep consistency with the input-output notation usually used in the GP literature.

tial coordinate \mathbf{x} corresponds to damage, is given by the probability that its novelty score exceeds the threshold T ; this can be obtained using the mean and variance estimates from the GP, which define a local Gaussian distribution for the novelty scores through \mathbf{m}_t and \mathbf{v}_t .

The probability that this uncertain measurement lies above the threshold, is thus given by the integral of the Gaussian probability density function,

$$p(y > T) = \int_T^{\infty} \mathcal{N}(\mathbf{m}_t, \mathbf{v}_t) dy \quad (2)$$

which can be evaluated from the cumulative density function of the Gaussian,

$$p(y > T) = \Phi\left(\frac{\mathbf{m}_t - T}{\mathbf{v}_t}\right) \quad (3)$$

This quantity is assessed at every spatial coordinate for which a GP prediction is available. When the variance of the GP is high, the probability of damage will be higher due to lack of information. Note that this measure is sensitive to the definition of T , so an estimate of the threshold should be used that truly captures the variability of the maxima of the novelty indices. The probability of damage here is the probability that the GP prediction exceeds the threshold set as the highest novelty index that would be attributed to a normal-condition process.

Fig. 4 provides an illustration of the problem using a small subset of the ultrasound NDT data from the wing panel. On the left, Fig. 4a shows a set of MSDs corresponding to the training data. Note that no values above the threshold are shown because, as per Algorithm 2 and the above discussion, these observations are discarded for the purpose of building a reference covariance model. On the right, Fig. 4b shows four different predictions of a GP, corresponding to 1) a low mean with low variance, 2) a low mean with high variance, 3) a high mean with high variance and 4) a high mean with low variance. Overlaid on the probability densities is the probability of damage, computed using Eq. (3).

This figure provides an intuitive view of this quantity; it is simply the amount of total probability mass above the threshold. In the case of an inlier GP prediction with low variance, this quantity is close to zero, and one can be confident that the observation is not an outlier. If variance increases, a more significant amount of probability mass lies above the threshold, and hence the probability of damage rises, even though the mean may still be within the threshold. In the case of outlying means, a high variance will tend to lower the probability of damage, while a low variance will have the effect of moving it closer to unity.

Note that the GP prediction is characterised by a Gaussian distribution; its tails decay to infinity, so there will always be some small probability mass either under or over the threshold, therefore never allowing a probability of damage of zero or unity.

4.2. Two-dimensional example

An illustration of the autonomous inspection procedure on a two-dimensional dataset is shown in Fig. 5 (a subset of the data shown in Fig. 2). As before, ToF and ultrasound signal attenuation were used as features. The section shown in Fig. 5 is the region around the delaminated area of the composite, on the top section.

The figure shows the evolution of the location of the observations as data is gathered, illustrating the mean and variance of the GP, and the probability of damage derived from them. Note that as soon as a region of high novelty values is discovered, the search prioritises exploitation around this region, until the uncertainty around it is significantly reduced, and it

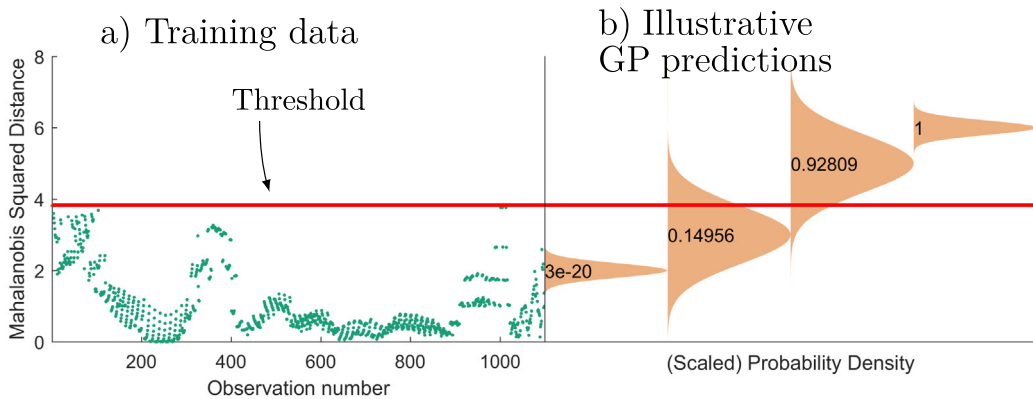


Fig. 4. Illustration of probability of damage, illustrating a) threshold placed over MSDs from a training data set and b) four illustrative GP predictions, two inliers and two outliers, each with low and high variance.

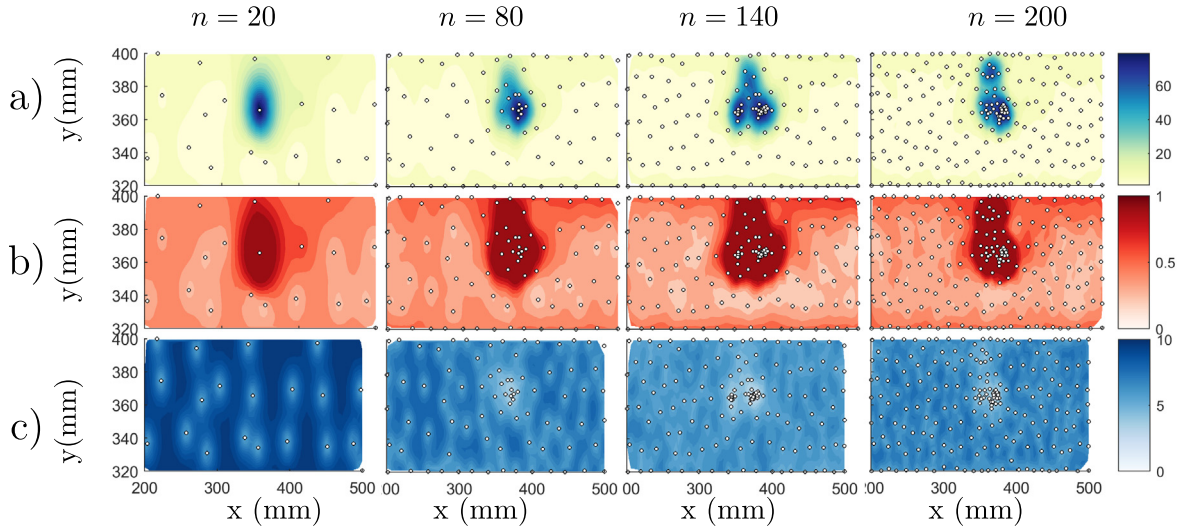


Fig. 5. a) Novelty scores (contour indicates GP mean while round markers denote measured locations), b) probability of damage, and c) GP standard deviation, from a two-dimensional subset NDT data from an aerospace specimen. Illustration shows evolution of predictions as observations are collected. n denotes the number of observations. The resulting probability of damage is shown inside each of the four Gaussians.

becomes more worthwhile to explore elsewhere. Note that the result of the search is close to a grid search in regions of low novelty, with a resolution given by ℓ , while a higher resolution of observations is placed around potentially-damaged regions.

5. Full scanning procedure results

Specimens containing different types of realistic damage are rare, so to demonstrate the full algorithm across the whole specimen, it is divided into eight regions with and without damage. These regions are shown in Fig. 6. Note that only regions

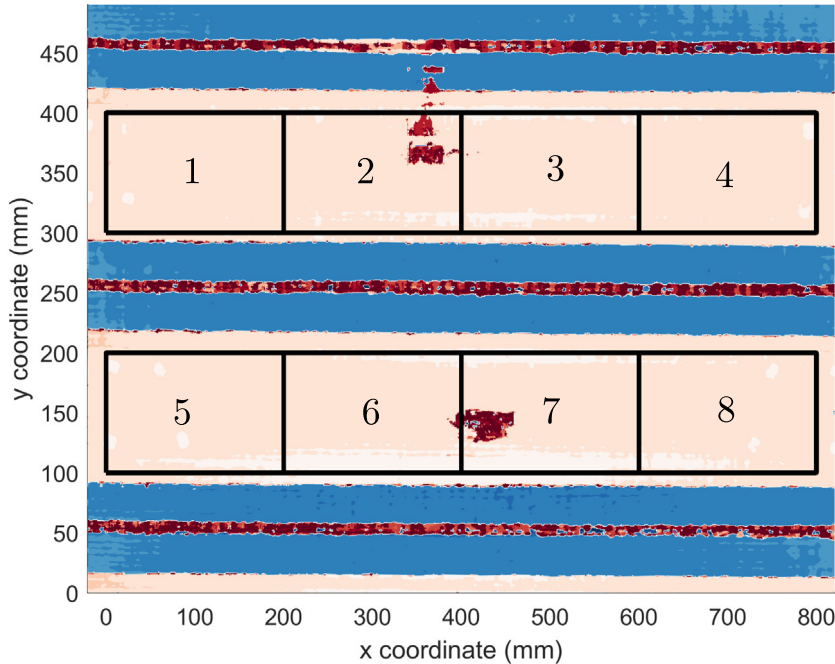


Fig. 6. Divisions of composite specimen used for experimental investigation. Colour-map indicates ultrasonic ToF. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

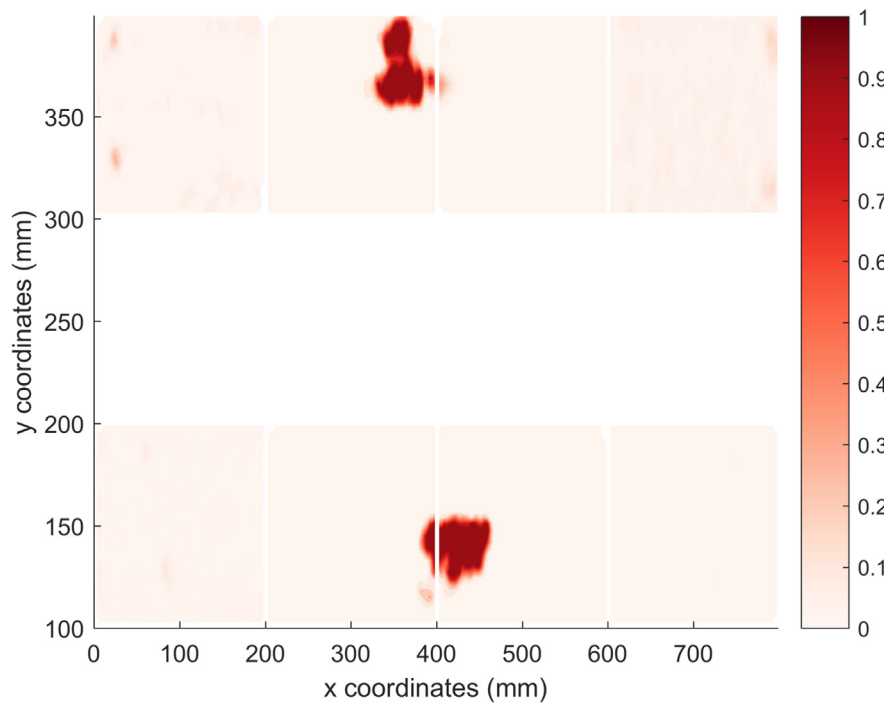


Fig. 7. Probability of damage contours resulting from the automated inspection algorithm, applied to eight local regions in the composite specimen.

from nominally-similar areas of the composite specimen are considered here. These are the thin wall sections. There are two areas of known delamination in this region, as well as areas free of damage.

The probability of damage resulting from the application of the autonomous inspection algorithm is shown in Fig. 7, as contours overlaid over the spatial coordinates of each of the eight sub-sections analysed individually. The damage probability contours correctly localise the two known areas of delamination with a near 100% probability of damage. The two delaminated regions have cross-sections of no less than 50 mm. As such, the result is not too surprising, given that a length-scale of $\ell = 10$ mm was used for this search, so any regions with observations at a spacing much more than ℓ will develop high uncertainty, and thus be prioritised by the search. This makes the identification of damage in regions two and six almost guaranteed. More interesting is the fact that the small area of delamination that was caught up in zone three has been detected with high probability, even though the maximum size of the delaminated area, (as observed through the original high resolution C-scan) has an overall maximum length of 5 mm.

As observations are collected sequentially, one is interested in the evolution of the probability of damage for every region. This is shown in Fig. 8. For each region of inspection, the interesting metric is the overall maximum observed damage; in order to summarise this, the 99th percentile of the estimated damage probability per region is plotted in Fig. 8. Note that this 'test' was run with a maximum budget of 500 observations per region, although some terminated early, due to the lowest predicted GP variance reaching the estimated σ_n^2 . In zones two and seven, where a probability of damage of 100% has been estimated, termination occurred as the algorithm attempted to collect data at a previously observed location, of high novelty index value. This is an artefact of applying the algorithm to pre-recorded data, as there is a finite discretisation grid. The only region that did not converge in this case is region eight, albeit with a very small probability of damage, less than machine precision.

During the first 100 observations, the damage probability tends to jitter; this is due to the fact that the robust statistics were re-estimated throughout the first 100 observations, and held fixed afterwards. Also, in regions with small manufacturing imperfections (evident as small round features, predominantly in zones one and four), the 99th percentile of the damage probability tends to jitter, instead of converging to a small value.

The regions with damage were found here with a very low number of observations. Any more observations from that point onwards only helped to confirm the presence, location and extent of damage (exploitation, in the Bayesian optimisation context), as well as exploring the remaining space, once the damage had been characterised well. Conversely, regions without damage took more observations to converge to a low damage probability, as it simply took more observations to reduce the uncertainty along all spatial coordinates by means of exploring the space. It is also interesting to see that in zone six, where only a 5 mm region of damage exists, the damage probability did not reach a level close to 100% until around 300 observations were acquired. Note that in this case there is a much larger spread in the resulting maxima of the damage probability, and it would take further observations to reduce this.

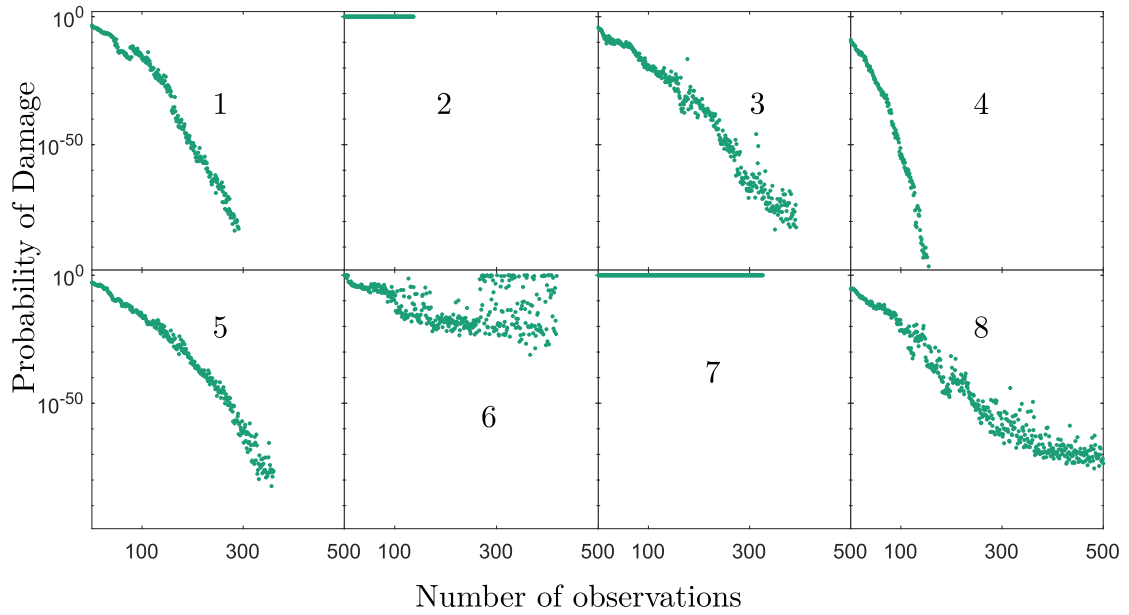


Fig. 8. Probability-of-damage evolution during sequential data acquisition, for the eight different regions analysed.

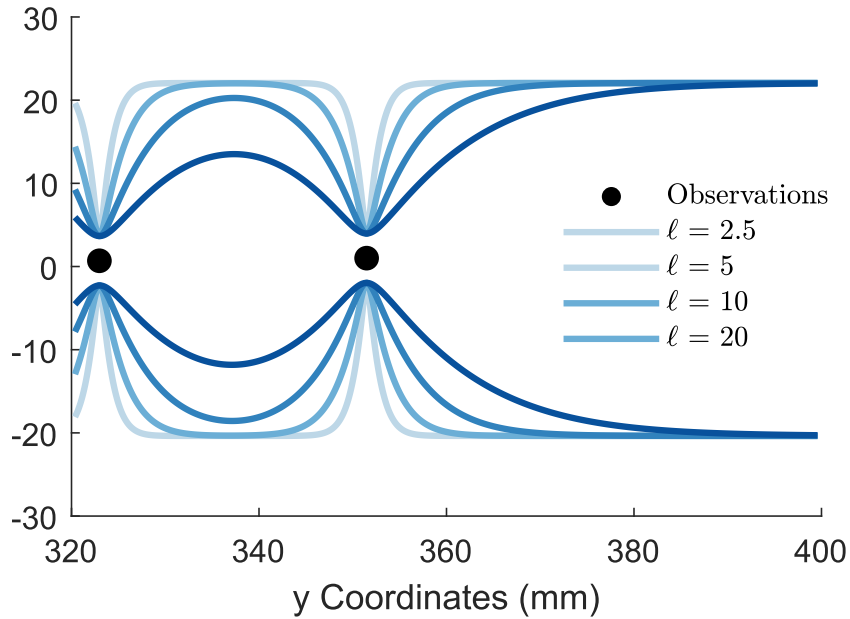


Fig. 9. Effect of length-scale hyperparameter on uncertainty predictions of the GP. Example shows 3σ bounds for four GP posteriors, conditioned on two observations, using increasing length-scales. A Matérn 3/2 covariance function was used.

The advantage of applying this autonomous inspection scheme is clear when one considers that the original C-scan data for each zone consisted of approximately 39,000 observations (individual ultrasound pulses, or A-scans). Performing the data collection sequentially using the Bayesian optimisation scheme reduces this significantly, while at the same time providing the required information in the form of a damage probability.

6. Discussion

A number of key factors affecting the algorithm implementation are discussed in this section.

6.1. Initialisation

A sensible implementation of the approach suggested here requires a number of initial scan points before optimisation can begin—this is a requirement from both the outlier analysis and the Bayesian optimisation.

Considering the robust estimation of the mean and covariance of the novelty indices, an issue arises from the dimensionality d of the features extracted from the data. The lower bound on the number of observations is twice the dimensionality of the feature vector, for the covariance to be well defined. From this point of view, at least $2d$ observations are required prior to commencing the autonomous scan. The issue is exacerbated by the fact that some of the observations may be potentially discarded from the set used in the FastMCD algorithm to compute the covariance.

A number of initial observations are also required to begin the Bayesian optimisation. For the GP to be contributing information, at least one observation is required. Even in the limit of such low observations, this *one shot* sequential design results in an initial GP model that is performing pure extrapolation at the initial stage of the algorithm. The GP predictions of where high function values and uncertain regions lie will tend to be poor in this extrapolation setting. The result would be an optimiser that takes much longer to converge, i.e. many more scans will be required.

It is beneficial to initialise the algorithm with a small number of points, by performing a low resolution scan. In the field of Optimal Experimental Design (OED) a popular choice is Latin Hypercube Sampling (LHS) [37]. The main feature of LHS is that it ensures that samples are well distributed over the space of each individual dimension. One of the drawbacks of LHS is that it tends to underestimate the variance of the objective function [38], this can have a serious impact on the performance of the objective function as a novelty detector. For this reason, random sampling from a uniform distribution, in the interval between the minimum and maximum x, y coordinates is used. This is sub-optimal in terms of space-filling, but is known to provide an unbiased estimate of objective function variance [38], which is ultimately more useful here.

Given the discussion above, the problem of exactly how many samples are required for initialisation translates to a problem of finding the number of samples n required to correctly capture the parameters of a Gaussian distribution of dimension p . The recommendation in [39] is to use a sample size of at least $14p$ in order to avoid non-convexities in the log-likelihood function used for estimation of the data feature covariance matrix. This specifies the minimum number of samples necessary to have a well-determined problem. The minimum number of samples required to correctly capture the variance of the novelty scores is a different one, and will in practice tend to be greater than $14p$. The important aspect is to gather enough samples to correctly capture the variance of the novelty indices, so in this paper, Monte Carlo samples are gathered until the novelty index variance converges to a stable value. In the wing panel ultrasound data used in this paper, in practice, 100 samples appeared to yield a stable variance.

6.2. Hyperparameter selection for GPs

Even though the GP is a 'nonparametric' regression technique [40], its behaviour is dependent on a small number of hyperparameters in the covariance function; the signal variance, length-scale and noise variance. Their values can greatly influence the result of the optimisation. It is essential that they are chosen in a sensible manner, since their values will influence the sensitivity of the proposed method. It has become commonplace in the GP literature to set the hyperparameters through an optimisation of the marginal likelihood of the GP [40]. However, it is also possible to set these hyperparameters based upon some prior knowledge about the function being modelled. In the case of the autonomous inspection problem this is the approach used.

The signal variance of the GP, σ_f , can control the exploration-exploitation balance of the optimiser. Because σ_f defines an upper limit on \mathbf{v}_t , if it is set too low, even with few observations, most areas of the space will default to σ_f^2 . The result is that the acquisition function (EI, or a similar measure) will choose to only scan at points of high mean values, thus performing pure exploitation; this would lead to a high resolution scan around areas where flaws are suspected (high \mathbf{m}_t), but an under-exploration of the rest of the space. In contrast, setting a high value of σ_f^2 would result in pure exploitation, as it would lead to extremely high uncertainty in the space observations. A middle ground is clearly necessary, and one reasonable solution would be to scale the data to unit standard deviation, and set the scaling hyperparameter to a reasonable value such as $\sigma_f^2 = 1$. However, the data in this case are novelty scores, evaluated sequentially, so the rescaling of the data would also need to happen sequentially, introducing another layer of complexity.

The length-scale of the covariance function ℓ , controls the rate in the input space at which the variance grows as one moves away from observed values, i.e. the rate of return to the prior. In the problem of spatial sampling this mediates the size of the region for which a sample can be said to be informative and as such is the most critical hyperparameter. A short length-scale implies that a sample only provides information about a very small region and vice versa. In this problem of spatial sampling, smaller length-scales will result in the GP variance growing at higher rates between observations far apart, while long length-scales will result in the GP predicting a lower variance between those same observations. This point is illustrated in Fig. 9 where the variance is shown for a GP conditioned on two observations for a range of growing length-scales.

In view of this, the length-scale can be seen as effectively controlling the risk one is willing to accept regarding detecting a specific flaw size. This is the tuning parameter of this algorithm that will have the largest impact on the resulting data acqui-

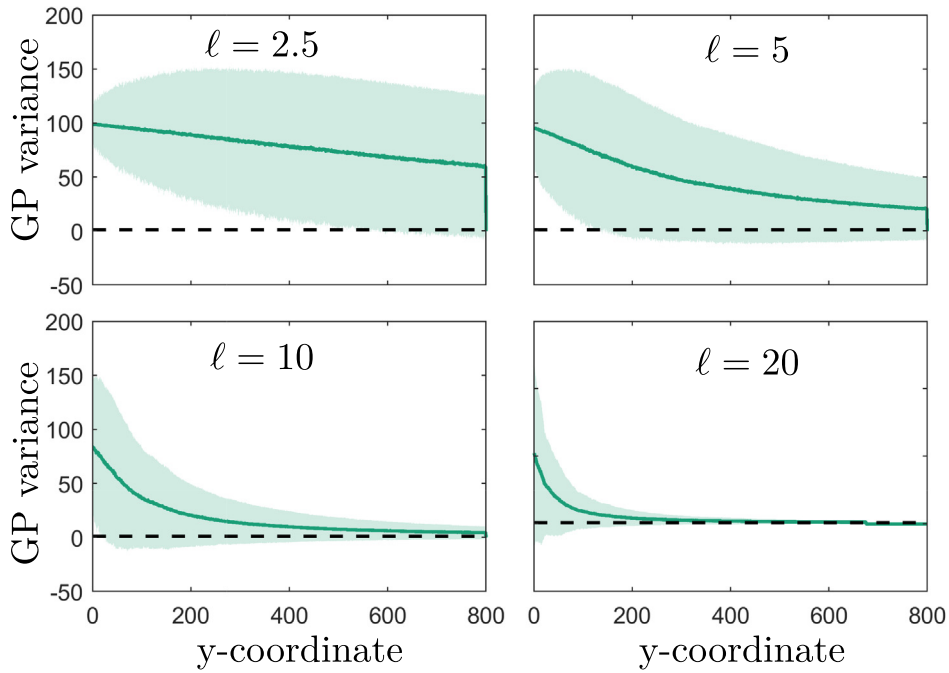


Fig. 10. Illustration of convergence rate with varying length-scales. The solid line represents the average variance of the GP predictions in two-dimensional space, the shaded area represents the 3σ bounds for the GP variance (the variance of the variance).

sition, and overall probability of damage estimates. The data acquisition will be altered through the effect of the length-scale on the acquisition function. Short length-scales will mean observations will be placed closer together, and vice versa.

The recommendation, in this work, is that the length-scale used should be based on the minimum expected flaw size. The form of the covariance function used can shed some light into what this implies. For instance, the (unscaled, $\sigma_f^2 = 1$) Matérn 3/2 function used in this paper, yields a correlation of $1/2$ at a distance ℓ apart, and a sharp decay in correlation beyond that. This means that good interpolation is expected from the GP at a distance of at least ℓ and placing observations closer to this is inefficient, whilst placing observations further away increases risk. In such a way the length scale of the process can be set *a priori* by the user on the basis of their requirements.

If a component were to be scanned that contained no anomalies whatsoever, this autonomous inspection scheme based on Bayesian optimisation should result in a scan with an average spatial resolution of ℓ (the acquisition of data would change when high novelty scores are found).

The other two hyperparameters here also play an important role in the uncertainty quantification given by the GP i.e. the process variance σ_f^2 and the noise variance σ_n^2 . The process variance will place an upper bound on the total possible uncertainty; it effectively specifies prior information on the uncertainty and defines the variance in the absence of any observations. The noise variance does exactly the opposite and places a lower bound on the uncertainty by specifying the variance around regions where many observations exist. The noise variance thus provides an indication of the inherent variability in novelty indices that is to be expected by taking the same observation multiple times at the same location.

6.3. Convergence

The lower bound on the variance of the GP prediction given by σ_n^2 is clearly important from the point of view of convergence criteria. If all the regions of high function values have been explored and the uncertainty across the whole range of possible measurement locations has been reduced to a level close to σ_n^2 , then there is little more benefit to be gained from further observations, and the search should be stopped. An estimate of the noise is available from the FastMCD.

The length-scale controls the total number of observations required to achieve a given spatial resolution (Section 6.2), shrinking the uncertainty across the specimen to the lower bound. Short length-scales lead to low convergence rates and vice versa. This effect is illustrated in Fig. 10, where the average variance and its 3σ bounds are shown for 800 iterations of Algorithm 2 over a two-dimensional sub-section of a composite specimen.

The final factor affecting when to stop scanning is the computational burden of the GP, this is not directly addressed in this paper although it is discussed (generally) in detail in the literature [41,42] and specifically for spatial data in [43].

7. Conclusions

A fully data-driven framework for autonomous inspection based on a Bayesian optimisation over robust novelty indices has been presented and demonstrated on an NDT data set consisting of ultrasound measurements on a high-value aerospace composite specimen. The specific specimen was used as a demonstration as it contains both damage-free and delaminated sections. The algorithm is data-driven, making use of robust inference of the mean and covariance of the NDT data features. This means that outliers that are present in the data set are identified and excluded from the estimation of the data statistics. The inspection scheme presented here is probabilistic, and it is aimed at minimising risk against an expected flaw size. The risk minimisation is provided by utilising Bayesian optimisation techniques, which also allow this algorithm to run sequentially. The impact is that the damage inference can be performed on-line and terminated either when no further information can be gained from the scan or when damage has been found. Bayesian optimisation is designed to find optimal points using the minimum number of observations possible. In the context of inspection, this implies that if a specimen does contain damage, or other anomalies, they will be found quickly, and the inspection will focus around these areas. If no damage is present the inspection will focus on minimising risk by means of exploring the space. This paper has focussed on the presentation of the algorithm. The key avenues of future work will be to address fully how this methodology can be implemented in real time and to demonstrate this on a full-system example. For this to happen it will be necessary to rigorously demonstrate the improvement in scanning speed over a pre-planned inspection path, to show how a computationally efficient form of the algorithm may be implemented and to demonstrate the effectiveness of physical selection of the hyperparameters of the algorithm. Overall, this algorithm enables autonomous inspection and it removes the human user decision-making regarding appropriate areas to scan, and instead replaces it with sequential decision making guided by on-line autonomous interpretation of the measured data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been funded by the UK Engineering and Physical Sciences Research Council (EPSRC) through grants EP/N018427/1, EP/S001565/1, EP/R003645/1 and EP/R004900/1. The authors would also like to acknowledge Dr. Charles MacLeod for his contribution to the experimental setup used for the ultrasound data gathering.

Appendix A. Robust outlier analysis

Outlier analysis often involves the evaluation of a distance metric that measures how far away a given observation lies from the centre of the data mass, relative to its spread. In the case of Gaussian-distributed data, the relevant metric is the Mahalanobis Squared-Distance (MSD),

$$d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) \quad (\text{A.1})$$

where the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are the two objects that characterise a multivariate Gaussian density. The observation vector \mathbf{y} represents the data features. Novelty detection with an MSD involves two steps: the first is training, and involves estimating the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from a training data set derived from a normal (undamaged) condition set. This step also involves the estimation of a novelty threshold for the novelty indices, T , above which observations are to be classed as novel. The second step is prediction, and involves evaluating a novelty index at new observations, and determining whether they fall above or below the novelty threshold. This type of analysis has been extensively used in data-driven Structural Health Monitoring (SHM) [3].

The top-level requirement in the data-driven autonomous inspection scheme is for the system to make a decision about whether the component contains outliers or not, based only on the data collected on that component alone. The implication of this is that outlying observations can potentially appear in the training set used for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This problem is well-suited for the application of *robust outlier analysis*, where robust measures of the mean and covariance of the data are estimated, which ignore outlying observations when estimating the Gaussian distribution parameters. Outlier analysis based on robust estimates of location and scatter in multivariate data has recently seen applications in the field of SHM [44], and the methodology used in this work closely follows the ideas of that reference.

A.1. Minimum covariance determinant estimators

There are two main approaches to robust estimation of Gaussian distribution parameters: the *Minimum Volume Ellipsoid* (MVE) [45,46] and the *Minimum Covariance Determinant* (MCD) [47,48]. MVE schemes seek to find an ellipsoid of minimum volume, which encloses at least half of the total number of data points; MCD schemes seek to find a covariance matrix $\boldsymbol{\Sigma}$ with

the smallest possible determinant that leaves out h observations from the training set, where h is the number of outliers. One advantage of the MCD over the MVE, is that the MCD is affine equivariant; location and scale parameters will remain constant under rescaling and translation of the data. Furthermore, the estimates of μ under MCD schemes have been shown to be asymptotically normal [49]. Beyond these theoretical guarantees, one useful practical advantage comes from the existence of an efficient way to compute MCD parameter estimates: the *Fast-MCD* algorithm [50]. Exact MCD estimates are numerically expensive to compute, requiring an evaluation of $n \times h$ subsets of size h . Fast-MCD uses an iterative scheme in order to achieve an MCD estimate with efficient computation, as described in Algorithm 3. The key step is to iteratively form h -subsets of the data matrix \mathbf{Y} , which contain values of low Mahalanobis distance (Eq. (A.1)), against the current estimates of Σ and μ .

Algorithm 3 Fast-MCD Algorithm

```

procedure Fast MCD ( $\mathbf{Y}, h$ )
  set  $\pi$  to be a random sample of  $h$  indices of  $\mathbf{Y}$ 
   $\mathbf{H}_1 \leftarrow \mathbf{Y}(\pi)$  ▷ form an  $h$ -subset of the data
  compute  $\mu_1, \Sigma_1$  from mean and covariance of  $\mathbf{H}_1$ 
  while  $\det(\Sigma_k) \neq \det(\Sigma_{k-1})$  do
     $d \leftarrow \text{mahal}(\mathbf{Y}, \mu_{k-1}, \Sigma_{k-1})$ 
    sort  $d$  in ascending order, store indices in  $\pi$ 
     $\pi \leftarrow \pi(1 : h)$  ▷ Discard  $n - h$  largest distances
     $\mathbf{H}_k \leftarrow \mathbf{Y}(\pi)$ 
    compute  $\mu_k, \Sigma_k$  from mean and covariance of  $\mathbf{H}_k$   $k \leftarrow k + 1$ 
  end while
  return  $\Sigma_k, \mu_k$ 
end procedure

```

While the Fast-MCD algorithm is naturally greedy, it is guaranteed never to increase the determinant of the covariance, $\det(\Sigma)$, at any iteration [50]; the stopping criteria for the algorithm is such that it continues as long as $\det(\Sigma_k) \neq \det(\Sigma_{k-1})$.

An application of covariance estimation with the Fast-MCD algorithm is given in Fig. A.11 for an illustrative synthetic data set. The data are samples from a two-dimensional Gaussian random variable, contaminated with outliers. The original distribution of the uncontaminated data is shown in light-grey, the maximum-likelihood covariance is plotted with a dashed line, while the Fast-MCD fit is shown with a solid line. The bias that the outliers introduce to the covariance is clear: they increase the scatter. The robust covariance gives a better approximation to the original density.

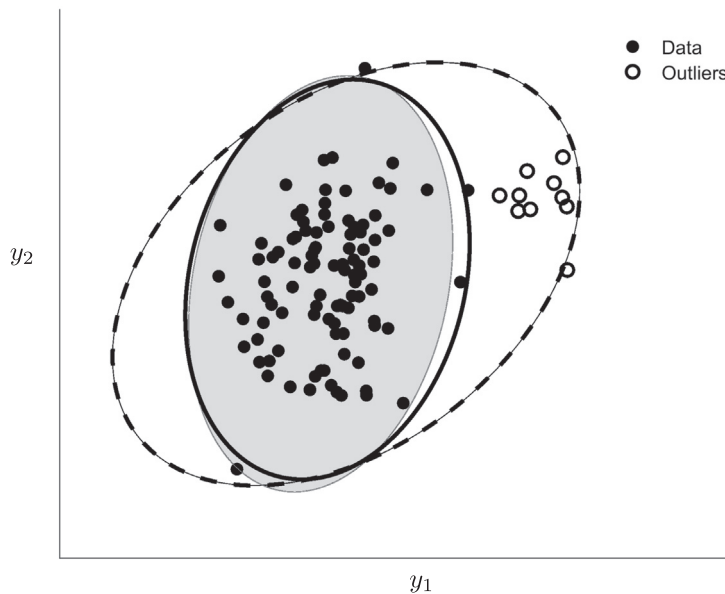


Fig. A.11. Illustration of robust estimation of Gaussian distribution parameters, in the presence of outliers.

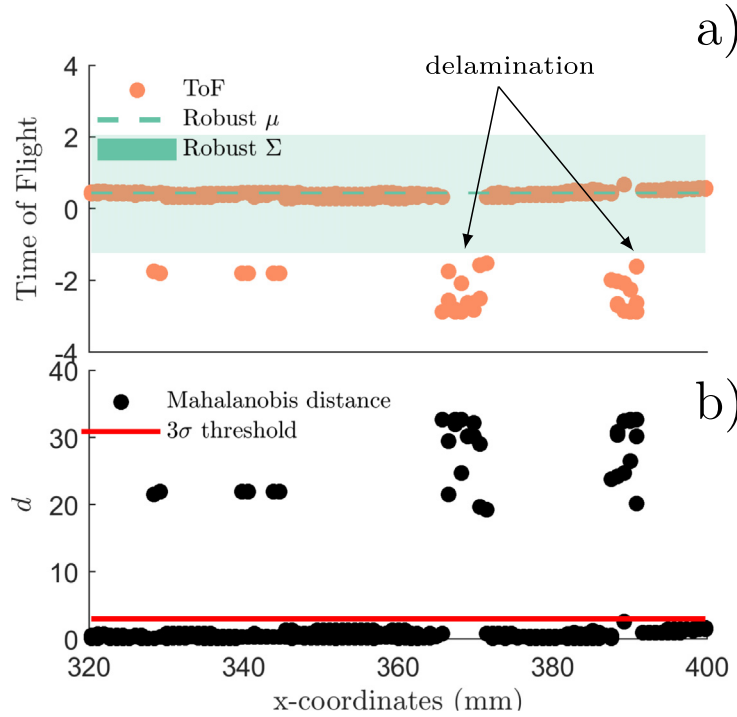


Fig. A.12. Illustration of outlier analysis on ultrasound-NDT data, using time-of-flight as a feature, and a robust mean and variance to compute a Mahalanobis distance.

A.2. Application to NDT data

The Mahalanobis-squared distance (MSD), using robust estimates of the data density, is central to the scanning procedure presented in this paper, which combines Bayesian optimisation with outlier analysis. The application of robust outlier analysis to NDT data will be briefly discussed here, in order to make clearer the application of Bayesian optimisation, presented in Section 2, to the problem of outlier analysis.

The goal is to assemble a mean and covariance of a reference, undamaged condition, even when there are outliers present in the data, which as discussed above, is often the case.

The NDT features discussed in Section 1.2 – namely, the Time of Flight (ToF) and attenuation – will be used here to demonstrate the idea of novelty detection, using robust measures. In order to simplify the problem, a one-dimensional task is considered; data from a cross section along the y-axis of the aerospace panel shown in Fig. 2 has been used.³ The ToF, together with the robust estimates (using Fast-MCD) for its mean and variance are shown in Fig. A.12a. This cross-section was selected as it contains two small regions where delamination of the carbon-fibre composite has occurred; these are evident as areas of much lower ToF compared to the rest of the specimen, due to the reflection of the wave from the delaminated section. Fig. A.12b shows the resulting robust MSD for all of the observations. The red horizontal line shows the threshold, which has been defined as the MSD of the 3σ values of \mathbf{Y} (just the ToF in this particular case). Any observations with an MSD above this threshold are defined as outlying or abnormal. The two regions where there is delamination have clear excursions above the threshold. Other outliers visible on the left side of the plot, are not at known damage locations, so it is likely that they have arisen from electrical noise or bad contact between the ultrasound probe and the specimen.

A.3. Threshold estimation

One important aspect of the algorithm, in particular the novelty detection part, is the definition of a threshold over the MSD above which observations are defined as outliers. There is more than one way of addressing the threshold selection problem; methods commonly used in SHM include the use of percentiles, Monte Carlo sampling [3], and extreme value statistics [51,52]. For simplicity here, the threshold was determined by computing the MSD of points six robust standard deviations away from the robust mean, along each dimension. This thresholding methodology makes use of the robust covariance computed by FastMCD, which has already been tuned for the separation of inclusive outliers.

Appendix B. Gaussian Process (GP) regression

This section discusses the estimation of the posterior density over the objective function given past observations, through the use of Gaussian Process (GP) regression. The use of GPs in the context of Bayesian optimisation dates back to [53], and later [54], as a better alternative to quadratic estimation of cost functions in multi-modal problems.

The Gaussian Process (GP) is a Bayesian nonlinear regression tool for modelling data which can be described in the following way,

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon \quad \varepsilon \sim \mathcal{N}(0, \sigma_n) \quad (\text{B.1})$$

where $\mathcal{N}(0, \sigma_n^2)$ denotes a Gaussian distribution with zero mean and variance σ_n^2 .

The GP defines a probability distribution over a function $f(\mathbf{x})$, using a mean $m(\mathbf{x})$, and a covariance function $k(\mathbf{x}, \mathbf{x}')$,

$$f(\mathbf{x}) = \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (\text{B.2})$$

Having observed a set of data $\mathcal{D}_t = \{X_t, \mathbf{y}_t\}$, predictions at a new test point are described by a mean \mathbf{m}_t and covariance \mathbf{v}_t [40],

$$\mathbf{m}_t = K(\mathbf{x}^*, X_t) \mathbf{P}^{-1} \mathbf{y}_t \quad (\text{B.3})$$

$$\mathbf{v}_t = K(\mathbf{x}^*, \mathbf{x}^*) - K(\mathbf{x}^*, X_t) \mathbf{P}^{-1} K(X_t, \mathbf{x}^*) \quad (\text{B.4})$$

where the matrix \mathbf{P} is defined as,

$$\mathbf{P} = K(X_t, X_t) + \sigma_n^2 \mathbf{I} \quad (\text{B.5})$$

B.1. Covariance functions

From the form of the model, clearly, the covariance function plays a critical role in controlling the behaviour of the GP. There are a number of choices for this function, some possible choices are discussed in [40]. In a stationary covariance function, the covariance is described as a function of the pairwise distance between observations,

$$\mathbf{r} = |\mathbf{x} - \mathbf{x}'| \quad (\text{B.6})$$

A commonly employed stationary kernel for modelling nonlinear functions is the Matérn kernel with the smoothness parameter set at $\nu = \frac{3}{2}$ [25,40], this is given by,

$$k_{\text{Mat}3/2}(\mathbf{r}) = \sigma_f^2 \left(1 + \frac{\sqrt{3}\mathbf{r}}{\ell} \right) \exp \left(-\frac{\sqrt{3}\mathbf{r}}{\ell} \right) \quad (\text{B.7})$$

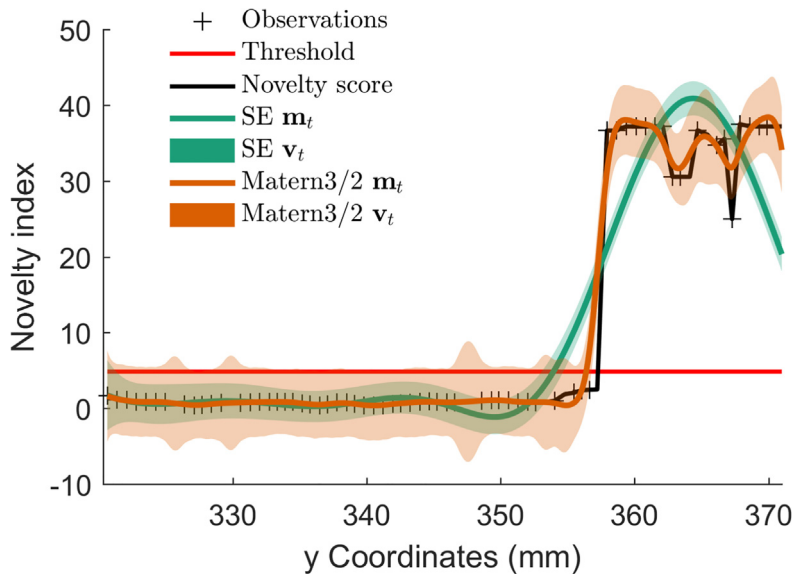


Fig. B.13. Illustration of GP posterior fit to novelty scores across a one-dimensional slice of ultrasound NDT data, comparing the use of squared-exponential and Matérn 3/2 covariance functions.

Fig. B.13 illustrates GP predictions with the same data set as used in the illustration in Fig. A.11. For comparison purposes, the GP predictions are shown using two different covariance functions; the Matérn 3/2 of Eq. (B.7) and the squared-exponential kernel [40]. The squared-exponential is known to produce smooth functions, since it is infinitely differentiable. Fig. B.13 zooms in on a region around the boundary between an undamaged and damaged condition. The Matérn kernel deals better with this sudden change. The length-scales in this case were fixed to $\ell = 10$ and the noise variance was set at $\sigma_n^2 = 10$. In this example, the squared-exponential kernel imposes unrealistic assumptions on the data, whereas the Matérn 3/2 kernel is better able to model the transition between undamaged and damaged conditions.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ymssp.2020.106897>.

References

- [1] C. Mineo, C. MacLeod, M. Morozov, S.G. Pierce, R. Summan, T. Rodden, D. Kahani, J. Powell, P. McCubbin, C. McCubbin, G. Munro, S. Paton, D. Watson, Flexible integration of robotics, ultrasonics and metrology for the inspection of aerospace components, in: AIP Conference Proceedings, vol. 1806, AIP Publishing LLC, 2017, p. 020026..
- [2] C. Mineo, S.G. Pierce, P.I. Nicholson, I. Cooper, Robotic path planning for non-destructive testing – a custom Matlab toolbox approach, *Robot. Comput.-Integr. Manuf.* 37 (2016) 1–12.
- [3] K. Worden, G. Manson, and N.R.J. Fieller, Damage detection using outlier analysis, *Journal of Sound and Vibration*, 229:647–667, 2000..
- [4] R. Fuentes, On Bayesian Networks for Structural Health Monitoring (PhD thesis), The University of Sheffield, 2017.
- [5] M. Friedrich, W. Galbraith, G. Hayward, Autonomous mobile robots for ultrasonic NDE, in: Proceedings – IEEE Ultrasonics Symposium, vol. 1, IEEE, 2006, pp. 898–901..
- [6] R. Bogue, The role of robotics in non-destructive testing, *Ind. Robot: Int. J.* 37 (2010) 421–426.
- [7] J.T. Stetson, W. De Odorico, Robotic inspection of fiber reinforced composites using phased array UT, AIP Conference Proceedings, vol. 1581, American Institute of Physics, 2014, pp. 1889–1895.
- [8] R. Summan, G. Dobie, J.J. Hensman, S.G. Pierce, K. Worden, D.O. Thompson, D.E. Chimenti, A probabilistic approach to robotic NDE inspection, *AIP Conf. Proc.* 10 (2010) 1999–2006.
- [9] S.R. Doctor, T.E. Hall, L.D. Reid, SAFT – the evolution of a signal processing technology for ultrasonic testing, *NDT Int.* 19 (1986) 163–167.
- [10] R. Drai, F. Sellidj, M. Khelil, A. Benchaala, Elaboration of some signal processing algorithms in ultrasonic techniques: application to materials NDT, *Ultrasonics* 38 (2000) 503–507.
- [11] F. Honarvar, H. Sheikhzadeh, M. Moles, A.N. Sinclair, Improving the time-resolution and signal-to-noise ratio of ultrasonic NDE signals, *Ultrasonics* 41 (2004) 755–763.
- [12] L. Udpa, S.S. Udpa, Neural networks for the classification of nondestructive evaluation signals, *IEE Proceedings F (Radar and Signal Processing)* 138 (1) (1991) 41–45.
- [13] H. Sohn, D. Dutta, J.Y. Yang, M. DeSimio, S. Olson, E. Swenson, Automated detection of delamination and disbond from wavefield images obtained using a scanning laser vibrometer, *Smart Mater. Struct.* 20 (2011) 045017.
- [14] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic NDT using artificial intelligence, *J. Nondestr. Eval.* 30 (2011) 20–28.
- [15] N.A. Akram, D. Isa, R. Rajkumar, L.H. Lee, Active incremental support vector machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers, *Ultrasonics* 54 (2014) 1534–1544.
- [16] T.M. Nunes, V.H.C. de Albuquerque, J.P. Papa, C.C. Silva, P.G. Normando, E.P. Moura, J.M.R.S. Tavares, Automatic microstructural characterization and classification using artificial intelligence techniques on ultrasound signals, *Expert Syst. Appl.* 40 (2013) 3096–3105.
- [17] R. Marchant, F. Ramos, Bayesian optimisation for informative continuous path planning, in: Proceedings – IEEE International Conference on Robotics and Automation, IEEE, 2014, pp. 6136–6143.
- [18] G. Cardoso, J. Saníe, Data compression and noise suppression of ultrasonic NDE signals using wavelets, in: IEEE Symposium on Ultrasonics, 2003, vol. 16, IEEE, 2015, pp. 250–253..
- [19] G.M. Zhang, C.Z. Zhang, D.M. Harvey, Sparse signal representation and its applications in ultrasonic NDE, *Ultrasonics* 52 (2012) 351–363.
- [20] R. Fuentes, C. Mineo, S.G. Pierce, K. Worden, E.J. Cross, A probabilistic compressive sensing framework with applications to ultrasound signal processing, *Mechanical Systems and Signal Processing* 117 (2019) 383–402.
- [21] J.-F. Wang, Ateín, B.-B. Gao, Y. Ge, A review of spatial sampling, *Spatial Stat.* 2 (2012) 1–14..
- [22] J. Jin, A.D. Heap, Spatial interpolation methods applied in the environmental sciences: A review, *Environ. Modell. Software* 53 (2014) 173–189.
- [23] N. Best, S. Richardson, A. Thomson, A comparison of Bayesian spatial models for disease mapping, *Stat. Methods Med. Res.* 14 (2005) 35–59.
- [24] F. Schoefs, E. Bastidas-Arteaga, T.V. Tran, G. Villain, X. Derober, Characterization of random fields from NDT measurements: a two stages procedure, *Eng. Struct.* 111 (2016) 312–322.
- [25] Michael L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Ser. Stat. (1999).
- [26] C. Gomez-Cardenas, Z.M. Sbartaí, J.P. Balayssac, V. Garnier, D. Breyse, New optimization algorithm for optimal spatial sampling during non-destructive testing of concrete structures, *Eng. Struct.* 88 (2015) 92–99.
- [27] H. Sang, A.E. Gelfand, Continuous spatial process models for spatial extreme values, *J. Agric. Biol. Environ. Stat.* 15 (2010) 49–65.
- [28] J.H. Kurz, A. Jüngert, S. Dugan, G. Dobmann, C. Boller, Reliability considerations of NDT by probability of detection (POD) determination using ultrasound phased array, *Eng. Fail. Anal.* 35 (2013) 609–617.
- [29] B.G.W. Yee, F.H. Chang, J.C. Couchman, G.H. Lemon, P.F. Packman, Assessment of NDE reliability data. Technical report..
- [30] F. Schoefs, A. Clément, A. Noud, Assessment of ROC curves for inspection of random fields, *Struct. Saf.* 31 (2009) 409–419.
- [31] Y. Humaida, P.D. Wilcox, M.D. Todd, B.W. Drinkwater, A probabilistic approach for the optimisation of ultrasonic array inspection techniques, *NDT&E Int.* 68 (2014) 43–52.
- [32] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, *J. Global Optim.* 13 (1998) 455–492.
- [33] S. Streltsov, P. Vakil, A non-myopic utility function for statistical global optimization algorithms, *J. Global Optim.* 14 (1999) 283–298.
- [34] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipiece curve in the presence of noise, *J. Basic Eng.* 86 (1964) 97.
- [35] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, *J. Global Optim.* 13 (1998) 455–492.
- [36] F. Mustapha, K. Worden, S.G. Pierce, G. Manson, Damage detection using stress waves and multivariate statistics: an experimental case study of an aircraft component, *Strain* 43 (2007) 47–53.
- [37] M.D. McKay, R.J. Beckman, W.J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (1979) 239–245.
- [38] M. Stein, Large sample properties of simulations using latin hypercube sampling, *Technometrics* 29 (1987) 143–151.

- [39] P. Zwiernik, C. Uhler, D. Richards, Maximum likelihood estimation for linear Gaussian covariance models, *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 79 (2017) 1269–1292.
- [40] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
- [41] J. Quinonero-Candela, C.E. Rasmussen, A unifying view of sparse approximate Gaussian process regression, *J. Mach. Learn. Res.* 6 (2005) 1939–1959.
- [42] T.D. Bui, J. Yan, R.E. Turner, A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation, *J. Mach. Learn. Res.* 18 (2017) 1–72.
- [43] C. Park, J.Z. Huang, Y. Ding, Domain decomposition approach for fast Gaussian process regression of large spatial data sets, *J. Mach. Learn. Res.* 12 (2011) 1697–1728.
- [44] N. Dervilis, E.J. Cross, R.J. Barthorpe, K. Worden, Robust methods of inclusive outlier analysis for structural health monitoring, *J. Sound Vib.* 333 (2014) 5181–5195.
- [45] P.J. Rousseeuw, Least median of squares regression, *J. Am. Stat. Assoc.* 79 (1984) 871–880.
- [46] D.M. Hawkins, J.S. Simonoff, High breakdown regression and multivariate estimation, *J. R. Stat. Soc. Ser. C* 42 (1993) 423–432.
- [47] D.M. Hawkins, The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data, *Comput. Stat. Data Anal.* 17 (1994) 197–210.
- [48] Mia Hubert, Michiel Debruyne, Minimum covariance determinant, *Wiley interdisciplinary reviews: Computational statistics*, Wiley Online Library 2 (1) (2010) 36–43.
- [49] R.W. Butler, P.L. Davies, M. Jhun, Asymptotics for the minimum covariance determinant estimator, *Ann. Stat.* 21 (1993) 1385–1400.
- [50] P.J. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics* 41 (1999) 212–223.
- [51] S.J. Roberts, Novelty detection using extreme value statistics, *IEE Proc. Vis. Image Signal Process.* 146 (1999) 124..
- [52] K. Worden, D.W. Allen, H. Sohn, D.W. Stinernes, C.R. Farrar, Extreme value statistics for damage detection in mechanical structures. Technical report, Los Alamos National Laboratory Report LA-13903-MS, 2002..
- [53] A. O'Hagan, J.F.C. Kingman, Curve fitting and optimal design for prediction, *J. R. Stat. Soc. Ser. B (Methodological)* 40 (1978) 1–42.
- [54] A. Zilinskas, On the use of statistical models of multimodal functions for the construction of the optimization algorithms, in: *Lecture Notes in Control and Information Sciences*, chapter 23, 1980..